Poster I-65

A Probabilistic-Based Approach in Modeling Protein Folding

Chen, Jie*[*1], White, Jason[1], Wang, Z. Jane[2], Liu, K.J. Ray[2]

[1]Division of Engineering, Brown University, Providence, RI, USA; [2]Department of Electrical and Computer Engineering/Institute for Systems Research, University of Maryland, College Park, MD, USA

One of the premier problems in quantitative biology is the *protein folding* problem of predicting the three-dimensional structure of a protein from its linear sequence of amino acids. The misfolding can lead to diseases such as Alzheimer's, Mad Cow, and Cystic Fibrosis diseases. Scientists have worked on the solutions ranging from numerical simulation of the physical forces exerted by the amino acids on one another to pattern recognition techniques which correlate motifs within the linear amino acid sequence with structural features of a protein. Although progress has been made by a variety of methods, this protein-folding problem is far from being solved.

In this paper, we develop a unified approach, where various techniques, such as belief propagation, Gibbs distribution, and CpG island methods have been employed, in providing an efficient way for modeling the protein folding. First, we model the DNA and CpG island transitions via Markov process. We propose to separate a protein sequence into different regions by discriminating between CpG island and non-CpG island regions. Second, by modeling protein folding as pairwise Markov random field (MRF), we propose a protein folding belief propagation (PF-BP) algorithm to compute the marginal probabilities, the *beliefs*, with linear computational cost. Third, we show the equivalence of the PF-BP to energy approximation. Overall, the protein folding problem is addressed by minimizing the Gibbs energy. Our motivation is that if all the local protein segments have folded into their optimal states (or minimum energy states), then these beliefs will propagate or lead to the belief that they will result in a global minimum energy folding state. Our proposed scheme forms a good cohesive connection between protein sequence data and the construction of digital signals. Simulation results are given to illustrate our proposed schemes. Our ongoing work is focusing on examining protein folding experiments to verify our claims.